

DOCUMENTACIÓ

PROJECTE DE PROGRAMACIÓ

Subgrup: 22

Clúster: 6.2

Alumnes:

Albert Ruiz Lombarte

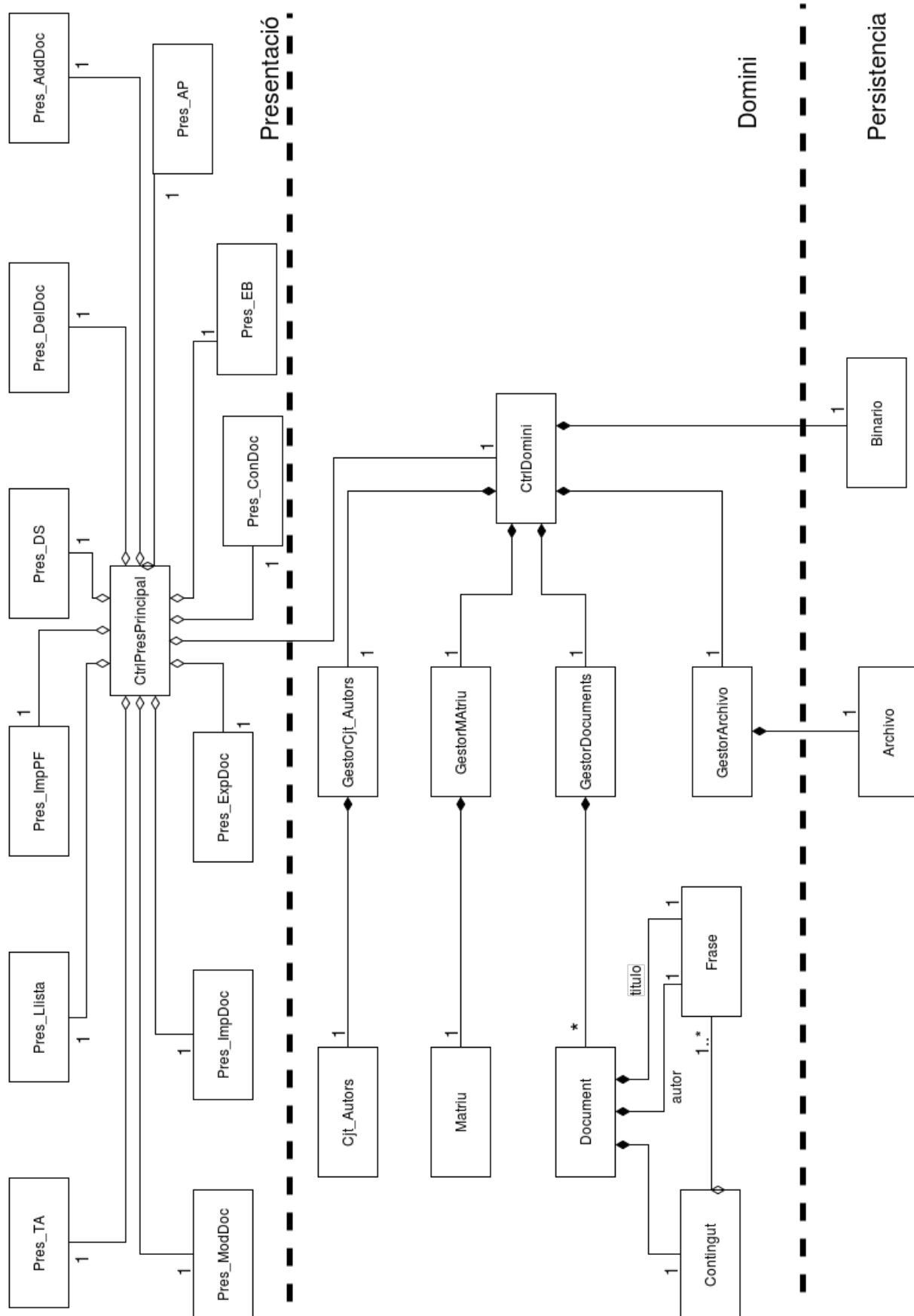
Aleix Roselló Rosiñol

John Bryan Game Glices

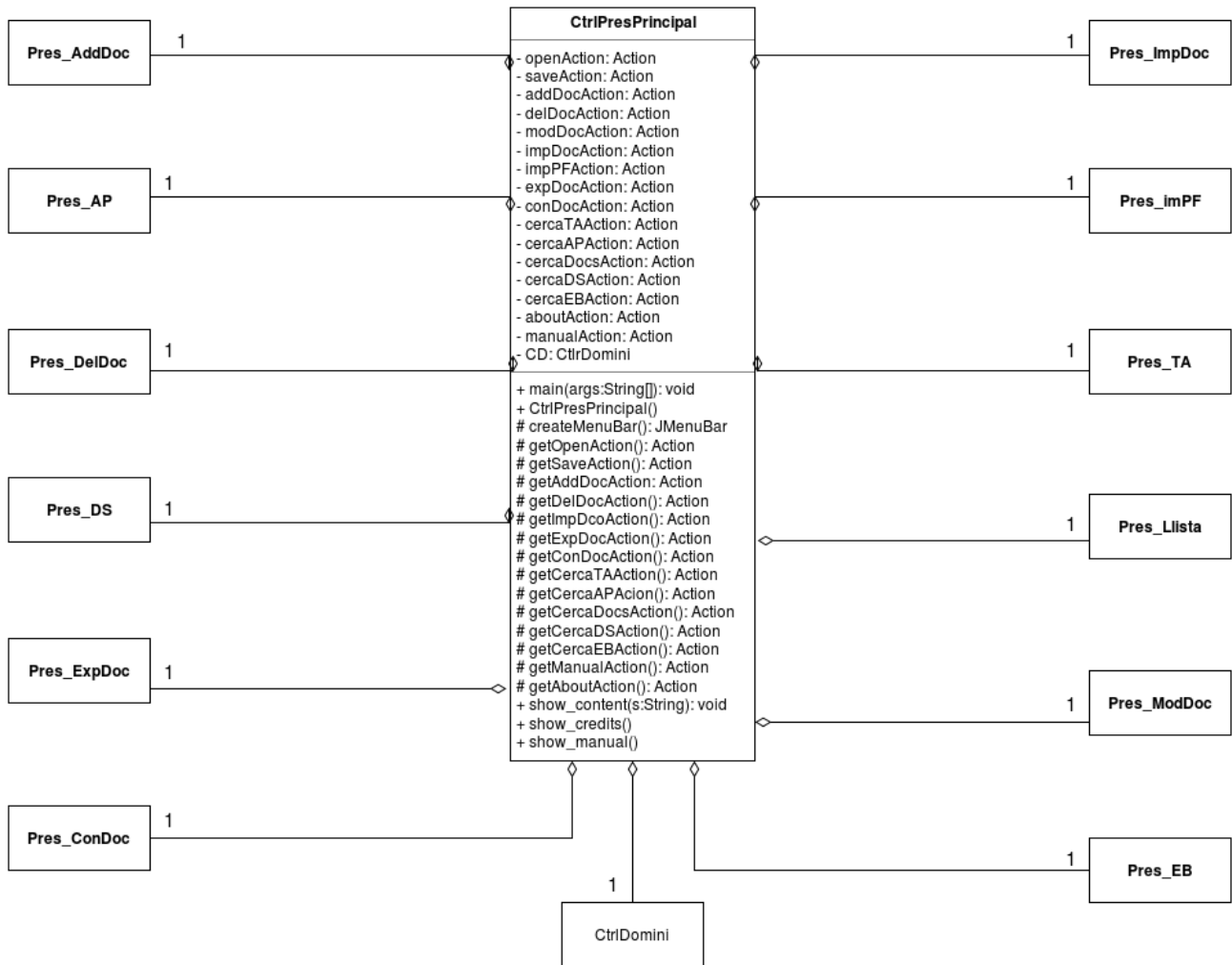
Índex

| | |
|---|-----------|
| 1. Disseny de l'arquitectura de tres capes (MVC)..... | 3 |
| 2. Diagrames UML estàtics complets generats per separat..... | 4 |
| 3. Descripció de les estructures de dades i algorismes..... | 20 |
| 4. Llista Classes Implementades..... | 23 |
| 5. Relació final de llibreries externes utilitzades..... | 24 |
| 6. Llistat de funcionalitats implementades..... | 25 |
| 7. Manual d'usuari..... | 29 |
| 8. Jocs de prova complets..... | 32 |

1. Disseny de l'arquitectura de tres capes (MVC)



2. Diagrames UML estàtics complets generats per separat



| Pres_impDoc |
|---|
| - i0: JLabel - i1: JLabel - i2: JLabel - t0: JTextField - t1: JTextField - fc: JFileChooser + CD: CtrlDomini + fr: JInternalFrame + pane: JDesktopPane |
| + Pres_impDoc(pane:JDesktopPane, fr:JInternalFrame, CD:CtrlDomini) - importar(titol:String, autor:String, arxiu:String): boolean - sortir(): void + actionPerformed(e:ActionEvent): void |

| Pres_impPF |
|--|
| - i0: JLabel - fc: JFileChooser + CD: CtrlDomini + fr: JInternalFrame + pane: JDesktopPane |
| + Pres_impPF(pane:JDesktopPane, fr:JInternalFrame, CD:CtrlDomini) - importarPF(arxiu:String): boolean - sortir(): void + actionPerformed(e:ActionEvent): void |

| Pres_TA |
|--|
| - i0: JLabel - t0: JTextField - btn: JButton + CD: CtrlDomini + fr: JInternalFrame + pane: JDesktopPane |
| + Pres_TA(pane:JDesktopPane, fr:JInternalFrame, CD: CtrlDomini) - sortir(): void - existeix_autor(autor:String): boolean - getTitols(autor:String): Object[] + actionPerformed(ev:ActionEvent): void + itemStateChanged(arg0:ItemEvent): void |

| Pres_Llista |
|--|
| + pane: JDesktopPane - l: JList |
| + Pres_Llista(pane:JDesktopPane, assumepte:String, array:Object[]) |

| Pres_ModDoc |
|--|
| - i0: JLabel - i1: JLabel - i2: JLabel - t0: JTextField - t1: JTextField - t2: JTextField - btn1: JButton - btn2: JButton - autor: String - titol: String + CD: CtrlDomini + fr: JInternalFrame + pane: JDesktopPane |
| + Pres_ModDoc(pane:JDesktopPane, fr:JInternalFrame, CD:CtrlDomini) - modificar(nou_con:String): void - get_con(): String + actionPerformed(ev:ActionEvent) + itemStateChanged(arg0:ItemEvent): void |

| Pres_EB |
|--|
| - i0: JLabel - t0: JTextField - btn: JButton + CD: CtrlDomini + fr: JInternalFrame + pane: JDesktopPane |
| + Pres_EB(pane:JDesktopPane, fr:JInternalFrame, CD:CtrlDomini) - getDocs(eb:String): Object[] + actionPerformed(ev:ActioEvent): void + itemStateChanged(arg0:ItemEvent): void |

| Pres_DelDoc |
|--|
| - I0: JLabel - I1: JLabel - t0: JTextField - t1: JTextField - btn: JButton + CD: CtrlDomini + fr: JInternalFrame + pane: JDesktopPane |
| + Pres_DelDoc(pane:JDesktopPane, fr:JInternalFrame, CD:CtrlDomini) - esborrar(): boolean - sortir(): void - actionPerformed(ev:ActionEvent): void + itemStateChanged(arg0:ItemEvent): void |

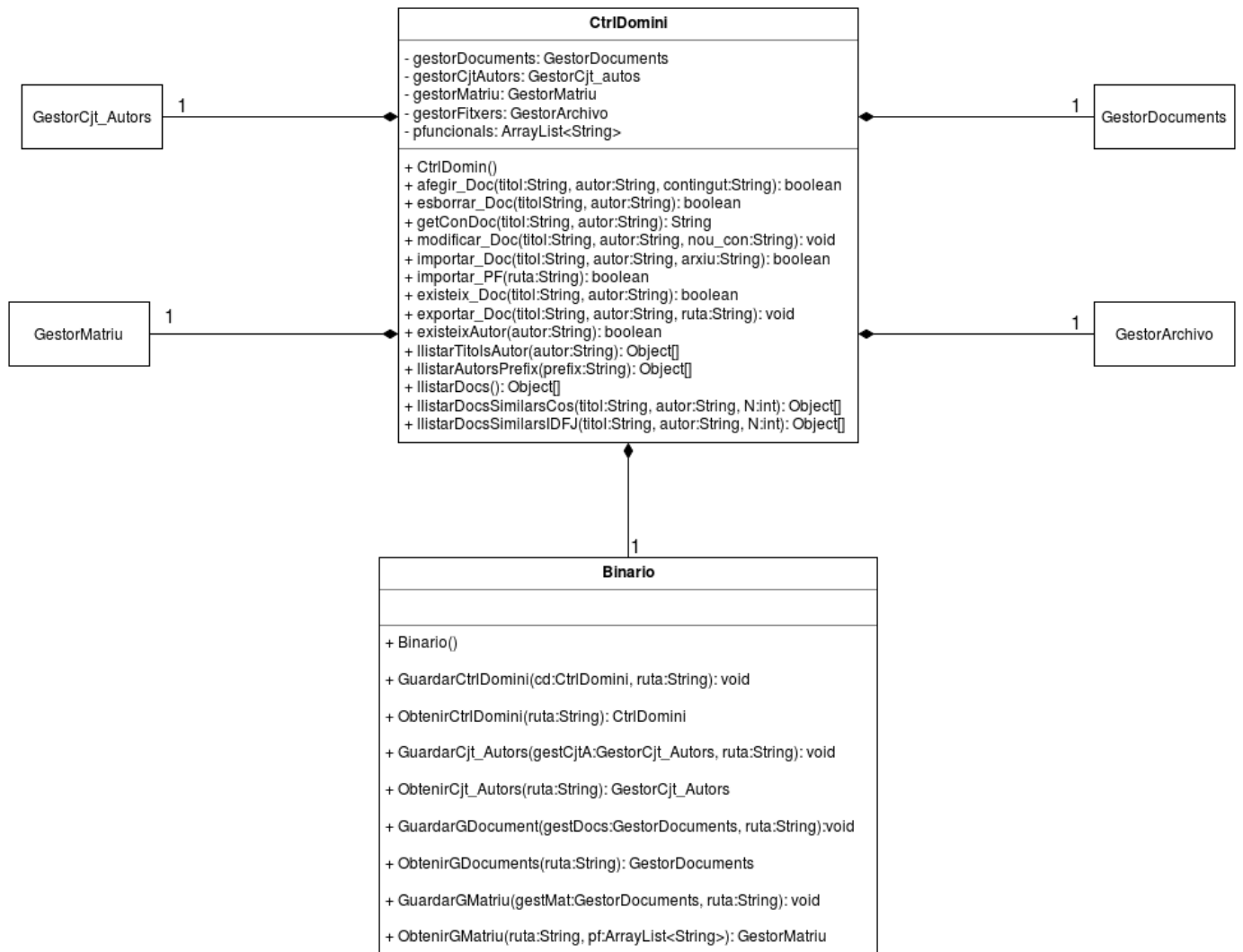
| Pres_ExpDoc |
|---|
| - I0: JLabel - I1: JLabel - I2: JLabel - t0: JTextField - t1: JTextField - fc: JFileChooser + CD: CtrlDomini + fr: JInternalFrame + pane: JDesktopPane |
| + Pres_ExpDoc(pane:JDesktopPane, fr:JInternalFrame, CD:CtrlDomini) - sortir(): void - existeixDoc(titol:String, autor:String): boolean - exportar(titol:String, autor:String, ruta:String): void + actionPerformed(e:ActionEvent): void |

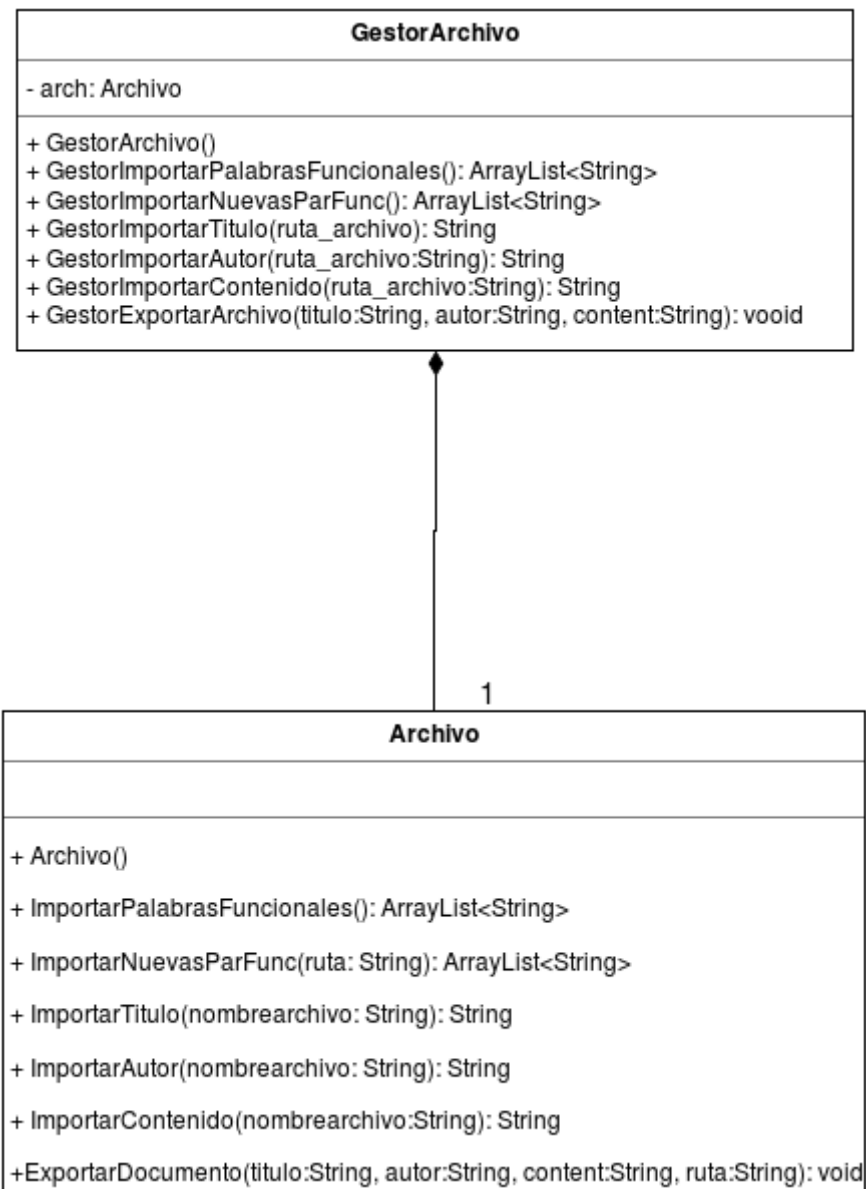
| Pres_DS |
|---|
| - I0: JLabel - I1: JLabel - I2: JLabel - I3: JLabel - t0: JTextField - t1: JTextField - t2: JTextField - ch: Choice - btn: JButton + CD: CtrlDomini + fr: JInternalFrame + pane: JDesktopPane |
| + Pres_DS(pane:JDesktopPane, fr:JInternalFrame, CD:CtrlDomini) - sortir(): void - existeix_Doc(titol:String, autor:String) - es_enter(n:String): boolean - getDocsCos(titol:String, autor:String, N:int): Object[] - getDocsIDFJ(titol:String, autor:String, N:int): Object[] - actionPerformed(ev:ActionEvent): void |

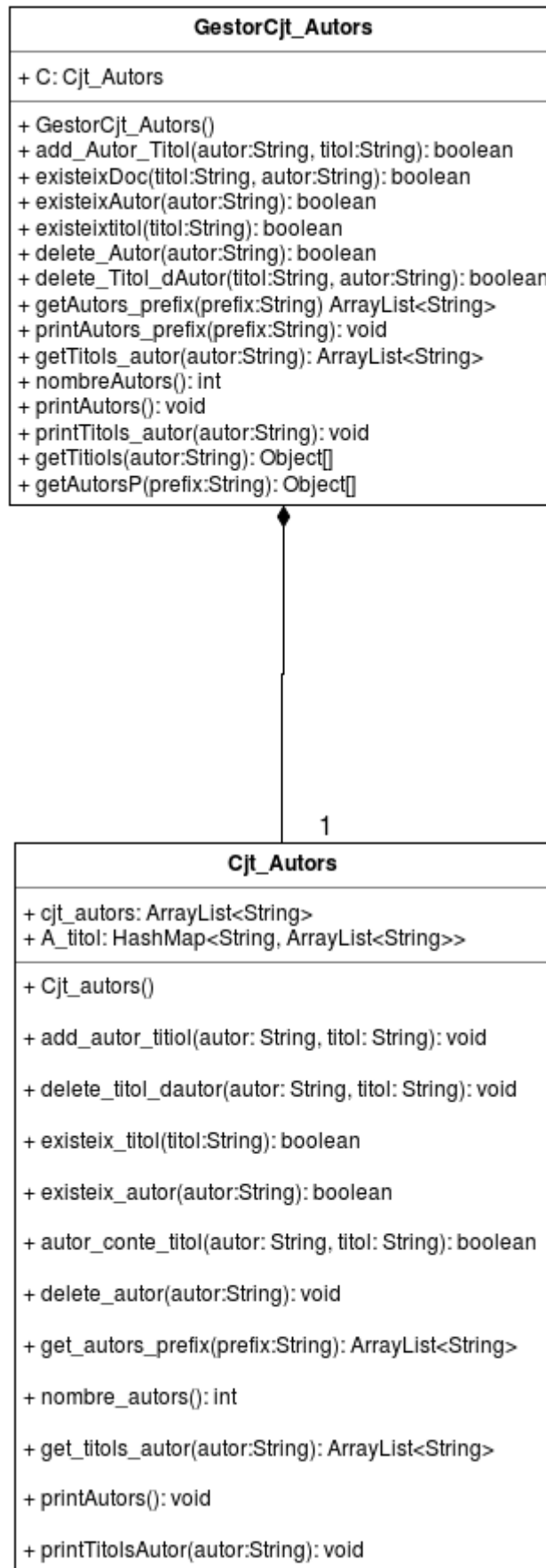
| Pres_ConDoc |
|--|
| - I0: JLabel - I1: JLabel - I2: JLabel - t0: JTextField - t1: JTextField - t2: JTextArea - btn1: JButton + CD: CtrlDomini + fr: JInternalFrame + pane: JDesktopPane |
| + Pres_ConDoc(pane:JDesktopPane, fr:JInternalFrame, CD:CtrlDomini) - get_con(titol:String, autor:String): String - sortir(): void + actionPerformed(ev:ActionEvent): void + itemStateChanged(arg0:ItemEvent): void |

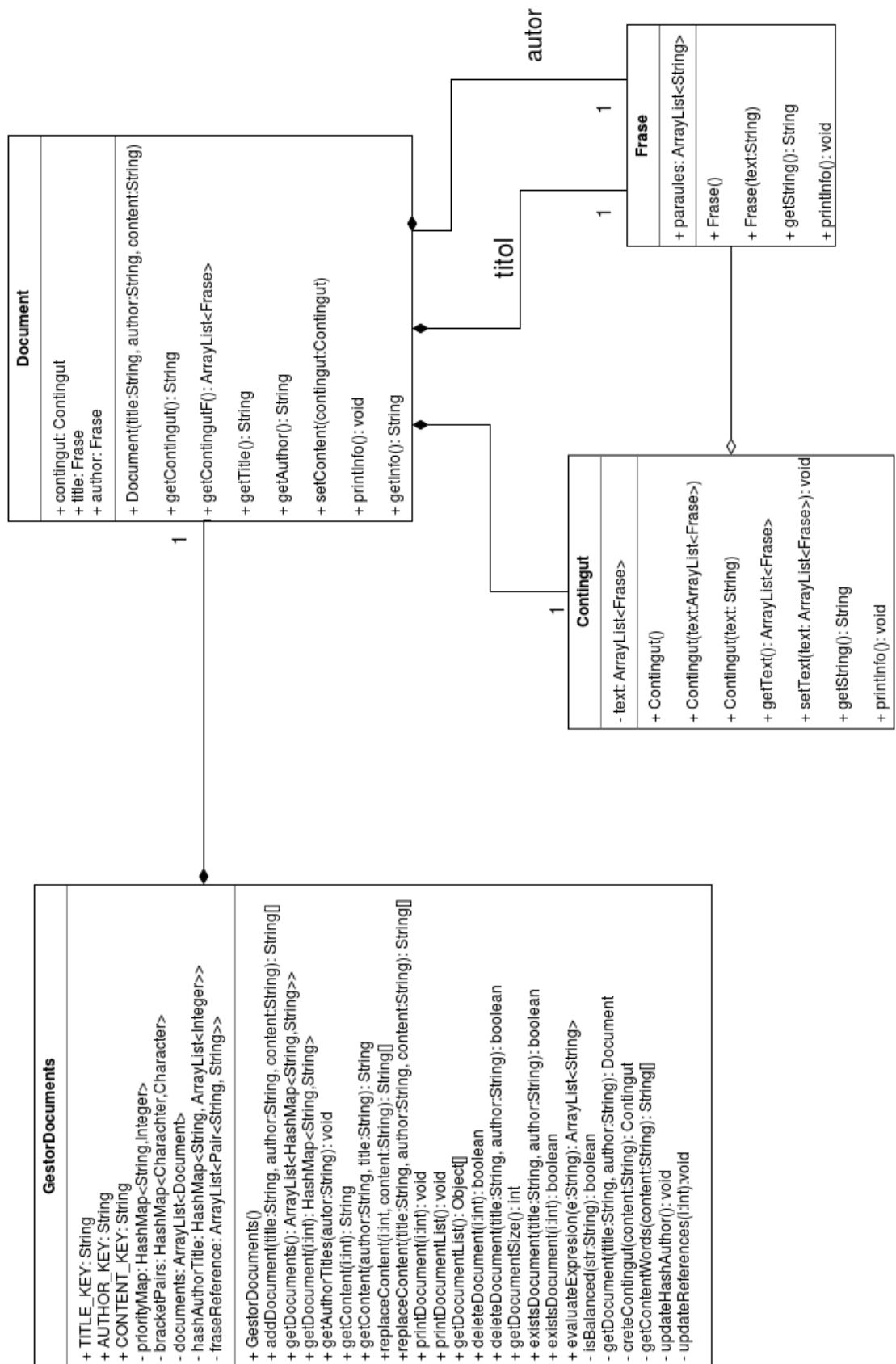
| Pres_AddDoc |
|---|
| <ul style="list-style-type: none"> - l0: JLabel - l1: JLabel - l2: JLabel - t0: JTextField - t1: JTextField - t2: JTextArea - btn: JButton + CD: CtrlDomini + fr: JInternalFrame + pane: JDesktopPane |
| <ul style="list-style-type: none"> + Pres_AddDoc(pane:JDesktopPane, fr:JInternalFrame, CD: CtrlDomini) - afegir(): boolean - sortir(): void + actionPerformed(ev:ActionEvent): void + itemStateChanged(arg0:ItemEvent): void |

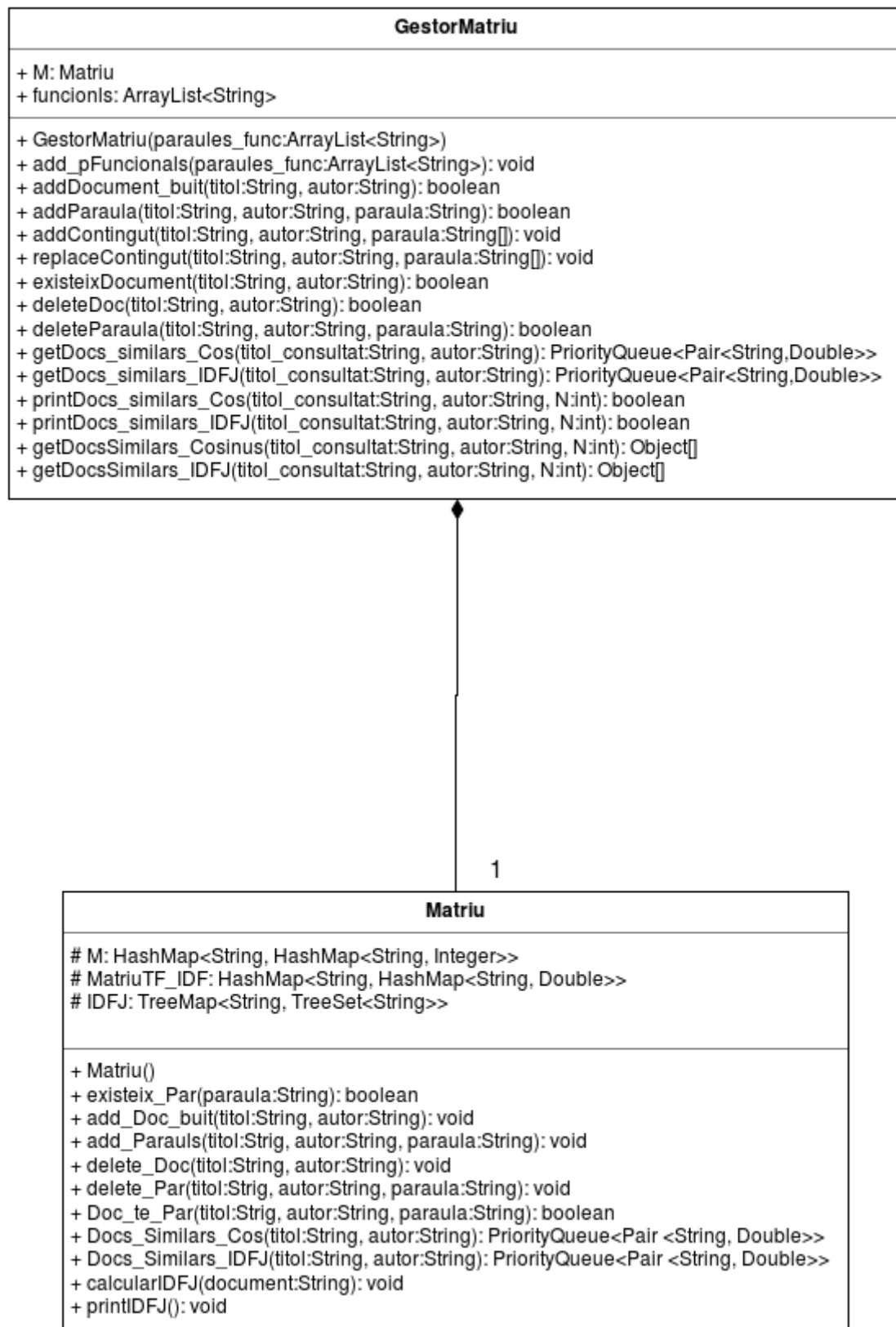
| Pres_AP |
|--|
| <ul style="list-style-type: none"> - l0: JLabel - t0: JTextField - btn: JButton - l: JList + CD: CtrlDomini + fr: JInternalFrame + pane: JDesktopPane |
| <ul style="list-style-type: none"> + Pres_AP(pane:JDesktopPane, fr:JInternalFrame, CD:CtrlDomini) - sortir(): void - getAutors(prefix:String): Object[] + actionPerformed(ev:ActionEvent): void + itemStateChanged(arg0:ItemEvent): void |











Descripció de les classes:

Nom: Archivo

Descripció: Arxiu creat per l'usuari per importar textos o exportar documents.

Atributs: No té atributs.

Nom: Binario

Descripció: Classe que permet a l'usuari obtenir i guardar intàncies d'una classe en format binari.

Atributs: No té atributs

Nom: Cjt_Autors

Descripció: Conjunt d'autors i de títols creat per l'usuari per agregar, obtenir o eliminar autors i títols.

Atribut1:

- Nom: cjt_autors

- Descripció: Conté els nom i cognom dels autors. Cada autor és únic.

- Estatic: No

Atribut2:

- Nom: A_titol

- Descripció: Per cada autor conté un conjunt de títols.

- Estatic: No

Nom: Contingut

Descripció: Contingut creat per l'usuari per modificar i obtenir continguts.

Atribut1:

- Nom: text

- Descripció: Conté un conjunt de frases.

- Estatic: Si

Atribut2:

- Nom: stringtext

- Descripció: Guarda el contingut ambs els salts de línia

- Estatic: Si

Nom: Document

Descripció: Document creat per l'usuari per modificar i obtenir documents

Atribut1:

- Nom: contingut

- Descripció: Contingut del document

- Estatic: No

Atribut2:

- Nom: title

- Descripció: Títol del document

- Estatic: No

Atribut3:

- Nom: author

- Descripció: Autor del document

- Estatic: No

Nom: Frase

Descripció: Frase creada per l'usuari per modificar i obtenir frases

Atribut:

- Nom: paraules
- Descripció: Paraules d'una frase.
- Estatic: No

Nom: Matriu

Descripció: Matriu creada per l'usuari per enmagatzemar una matriu esparsa amb nomïes els valor diferents de zero

Atribut1:

- Nom: M
- Descripció: Conté títols i autors i per qualsevol conté una paraula amb les seves aparicions.
- Estatic: No

Atribut2:

- Nom: MatriuTF_IDF
- Descripció: Matriu que té els pesos IDF
- Estatic: No

Atribut3:

- Nom: IDFJ
- Descripció: Matriu amb una paraula com a clau i el contingut son els documents que tenen aquesta paraula.
- Estatic: No

Nom: GestorArchivo

Descripció: Gestor d'arxiu creat per l'usuari per gestionar la classe Archivo

Atribut:

- Nom: arch
- Descripció: Arxiu d'un Gestor d'arxius
- Estatic: Si

Nom: GestorCjt_Autors

Descripció: Gestor de conjunt d'autors creat per l'usuari per gestionar la classe Cjt_Autors

Atribut:

- Nom: C
- Descripció: Conjunt d'autors d'un Gestor de conjunt d'autors
- Estatic: No

Nom: GestorDocuments

Descripció: Gestor de documents creat per l'usuari per gestionar la classe Document

Atribut1:

- Nom: TITLE_KEY
- Descripció: Clau d'un hashmap amb un contingut d'un títol
- Estatic: No

Atribut2:

- Nom: AUTHOR_KEY
- Descripció: Clau d'un hashmap amb un contingut d'un autor
- Estatic: No

Atribut3:

- Nom: CONTENT_KEY
- Descripció: Clau d'un hashmap amb un contingut d'un contingut
- Estàtic: No

Atribut4:

- Nom: documents
- Descripció: Conjunt de documents del Gestor de documents
- Estàtic: No

Atribut5:

- Nom: hashAuthorTitle
- Descripció: Conté autors i títols dels documents
- Estàtic: No

Atribut6:

- Nom: priorityMap
- Descripció: Conté els caràcters «(», «|», «!», «&»
- Estàtic: Si

Atribut7:

- Nom: bracketPairs
- Descripció: Conté una parella de cada tipus de parèntesis
- Estàtic: Si

Atribut8:

- Nom: fraseReference
- Descripció: Guarda les frases de tots els documents
- Estàtic: Si

Nom: GestorMatriu

Descripció: Gestor d'una matriu creat per l'usuari per gestionar la classe Matriu

Atribut1:

- Nom: M
- Descripció: Matriu d'un Gestor de matrius
- Estàtic: No

Atribut2:

- Nom: funcionals
- Descripció: Conté paraules funcionals
- Estàtic: No

Nom: CtrlDomini

Descripció: Classe que conté un conjunt de gestors que permet a l'usuari afegir, esborrar, obtenir, modificar, importar, exportar documents. També li permet fer una llista d'autors, títols i documents.

Atribut1:

- Nom: gestorDocuments
- Descripció: Gestiona els documents
- Estàtic: Si

Atribut2:

- Nom: gestorCjAutors
- Descripció: Gestiona un conjunt d'autors
- Estàtic: Si

Atribut3:

- Nom: gestorMatriu
- Descripció: Gestiona la comparacio dels documents
- Estàtic: Si

Atribut4:

- Nom: gestorFitxers
- Descripció: Gestiona l'importacio i exportacio de fitxers
- Estàtic: Si

Atribut5:

- Nom: pFuncionals
- Descripció: Conté les paraules funcionals
- Estàtic: Si

Nom: Pres_DelDoc

Descripció: Menú interactiu que permet a l'usuari esborrar un o més documents.

Nom: Pres_AddDoc

Descripció: Menú interactiu que permet a l'usuari afegir un o més documents.

Nom: Pres_DS

Descripció: Menú interactiu que permet a l'usuari llistar documents mitjançant el càlcul amb el mètode del cosinus o mitjançant el càlcul amb el mètode TF-IDFJ.

Nom: Pres_EB

Descripció: Menú interactiu que permet a l'usuari llistar el documents que compleixen l'expressió booleana.

Nom: Pres_ExpDoc

Descripció: Menú interactiu que permet a l'usuari exportar un o més documents.

Nom: Pres_ImpDoc

Descripció: Menú interactiu que permet a l'usuari importar un o més documents.

Nom: Pres_ImpPF

Descripció: Menú interactiu que permet a l'usuari importar un conjunt de paraules funcionals.

Nom: Pres_Llista

Descripció: Menú interactiu que permet a l'usuari_mostrar llistes de contingut divers depenent del que s'hagi volgut llistar.

Nom: Pres_ModDoc

Descripció: Menú interactiu que permet a l'usuari modificar un o més documents.

Nom: Pres_TA

Descripció: Menú interactiu que permet a l'usuari llistar el títols d'un autor.

Nom: Pres_AP

Descripció: Menu interactiu que permet a l'usuari llistar autors segons el prefix.

Nom: Pres_ConDoc

Descripció: Menu interactiu que permet a l'usuari mostrar el contingut d'un document.

Nom CtrlPresPrincipal:

Descripció: Menú interactiu que permet a l'usuari carregar i guardar sessions i afegir, esborrar, .obtenir, modificar, importar, exportar documents. També li permet fer una llista d'autors, títols i documents.

Atribut1:

-Nom: openAction

-Descripció: S'utilitza per iniciar una sessió

-Estàtic: Si

Atribut2:

-Nom: saveAction

-Descripció: S'utilitza per guardar una sessió

-Estàtic: Si

Atribut3:

-Nom: addDocAction

-Descripció: S'utilitza per afegir un document

-Estàtic: Si

Atribut4:

-Nom: delDocAction

-Descripció: S'utilitza per eliminar un document

-Estàtic: Si

Atribut5:

-Nom: modDocAction

-Descripció: S'utilitza per moidificar un document

-Estàtic: Si

Atribut6:

-Nom: impDocAction

-Descripció: S'utilitza per importar un document

-Estàtic: Si

Atribut7:

-Nom: impPFAction

-Descripció: S'utilitza per importar paraules funcionals

-Estàtic: Si

Atribut8:

-Nom: expDocAction

-Descripció: S'utilitza per ecportar un document

-Estàtic: Si

Atribut9:

- Nom: conDocAction
- Descripció: S'utilitza per consultar un document
- Estàtic: Si

Atribut10:

- Nom: cercaTAAction
- Descripció: S'utilitza per iniciar una sessió
- Estàtic: Si

Atribut11:

- Nom: openAction
- Descripció: S'utilitza per llistar títols d'un autor
- Estàtic: Si

Atribut12:

- Nom: openAction
- Descripció: S'utilitza per iniciar una sessió
- Estàtic: Si

Atribut13:

- Nom: cercaAPAction
- Descripció: S'utilitza per consultar autors per prefix
- Estàtic: Si

Atribut14:

- Nom: cercaDocsAction
- Descripció: S'utilitza per llistar els documents que es troben al sistema
- Estàtic: Si

Atribut15:

- Nom: cercaDSAction
- Descripció: S'utilitza per llistar documents similars
- Estàtic: Si

Atribut16:

- Nom: cercaEBAction
- Descripció: S'utilitza per llistar documents que compleixen una expressió booleana
- Estàtic: Si

3. Descripció de les estructures de dades i algorismes

Estructures de dades:

Classe Cjt_Autors

La classe Cjt_Autors conté dues estructures la primera consisteix en una llista dels noms dels autors que hi ha al sistema ordenats alfabèticament. La segona conté un HashMap on cada clau correspon a un autor i el valor de cada clau correspon a una llista, representada per un ArrayList, que conté els títols de les obres que ha escrit aquell autor ordenades alfabèticament.

Classe Matriu

La classe Matriu conté una matriu esparsa, és a dir una matriu on cada fila no conté el mateix nombre de columnes. Cada fila correspon a un document identificat pel seu títol i el seu autor, i dins de cada fila hi ha el conjunt de paraules no funcionals del seu contingut i amb el seu nombre d'aparicions. Per implementar-ho s'utilitza un HashMap on les seves claus, que serien les files de la matriu, són el títols i els autors que indiquen el document. Cada clau conté un altre HashMap les claus del qual són les paraules no funcionals del contingut d'aquell document, cada paraula conté un enter que correspon al nombre d'aparicions d'aquella paraula en aquell contingut.

Classe Pair

La classe Pair consisteix en la representació de la combinació de dos valors, el primer és l'identificador de un document, és a dir el seu títol i el seu autor, i el segon un nombre real que correspon al grau de semblança que té amb un altre document.

Classe Frase

La classe Frase conté una única estructura que es tracta d'una llista amb les paraules de conté la frase. Es tracta d'una ArrayList de string que pot incloure també signes de puntuació ja que la frase es separa per espais sense tenir en compte els signes de puntuació, ja que es tenen en compte abans.

Classe Contingut

La classe Contingut conté una única estructura que es tracta d'una llista amb les frases que conté. Aquesta llista es separa a través dels signes de puntuació, fins a cada punt es considera una frase. Està representat en forma d'ArrayList de Frase.

Classe Document

La classe Document conté tres estructures, dos de tipus frase, una representa el títol del document i l'altre l'autor de l'obra. Finalment conté una estructura del tipus contingut on s'emmagatzema el contingut del document seguint les especificacions de la classe contingut.

Classe ExpresionTree

La classe ExpresionTree conté dos estructures, una de tipus HashMap que conté una llista de prioritats de les operacions d'una expressió booleana i un node que representa l'arrel de l'arbre representat per l'expressió.

Classe Node

La classe Node conté tres estructures, dos del mateix tipus Node representant els fills esquerra i dret (si en té) i una altra variable que conté el valor del node.

Explicació dels algorismes:

Obtenció dels títols de les obres que ha escrit un autor:

Per obtenir els títols que ha escrit un determinat autor es retorna la llista de títols ordenats alfabèticament a partir de la clau del HashMap que correspon al nom de l'autor.

Obtenció dels autors segons els seu prefix:

Per obtenir els autors a partir d'un prefix es realitza una cerca dicotòmica amb el prefix indicat per l'usuari sobre la llista d'autors ordenats alfabèticament, un cop sabuda la posició on es trobaria el prefix respecte la llista és comprova que el nom de l'autor comença per aquell prefix, així es va avançant fins que hi ha un que no ho compleix o s'ha arribat al final de la llista.

Obtenció del contingut donat el seu títol i autor:

Per obtenir el contingut d'un document donat el seu títol i autor, primer es comprova que l'autor existeixi en el HashMap del gestor, on es troben tots els autors com a clau i el valor es tracta d'una llista amb les posicions on l'autor té un document. A partir d'aquesta llista es recorre buscant el títol que correspongui amb el donat per paràmetre i aleshores es retorna el contingut del document.

Obtenció dels títols que compleixen una expressió booleana:

Per obtenir la llista de documents que compleixen una expressió booleana el primer que s'ha de tenir en compte és que l'expressió ve donada amb el tipus d'inordre d'un arbre, i per tal de construir l'arbre que representa aquesta expressió cal convertir la expressió a post-ordre. També s'ha de comprovar que l'expressió està ben parentitzada, i per tant es comprova que per cada parèntesi o clau d'obertura n'hi ha una que la tanca mitjançant una pila. Un cop convertida, cal construir l'arbre que representa l'expressió, on tan operacions com continguts estan representats per un node. A partir de l'arbre, li passem totes les frases de tots els continguts del programa, i les anem eliminant del conjunt a mesura que no van complint les diferents sub-parts de l'expressió, fins obtenir finalment només el conjunt resultant de frases que la compleixen.

Càlcul de semblances entre documents amb el mètode Cosinus:

Per calcular les semblances entre dos documents, es representen com a vectors el conjunt de paraules no funcionals de manera que queden de la següent forma:

document1 = (pes_del_terme 1, pes _del_terme 2, ..., pes_del_terme n)
document2 = (pes_del_terme 1, pes _del_terme 2, ..., pes_del_terme n)

·
·
·

documentN = (pes_del_terme 1, pes _del_terme 2, ..., pes_del_terme n)

document_consulta = (pes_del_terme 1, pes _del_terme 2, ..., pes_del_terme n)

On cada pes de terme correspon al nombre d'aparicions que té cada paraula no funcional en el contingut del document. Per determinar la similitud entre dos documents es calcula el cosinus de l'angle que formen aplicant el Teorema del Cosinus, és a dir amb la següent fórmula:

$$\cos(\vec{v}, \vec{w}) = \frac{\vec{v} \times \vec{w}}{\|\vec{v}\| \cdot \|\vec{w}\|}$$

On \vec{v} i \vec{w} representen dos documents, $\vec{v} \times \vec{w}$ és el producte escalar dels dos vectors i $\|\vec{v}\| \cdot \|\vec{w}\|$ és el producte dels seus mòduls. El resultat dóna un nombre entre 0 i 1 que representa el grau de similitud que té un document amb l'altre, com més proper a 1 més semblants seran els documents i com més proper a 0 més diferents.

Càlcul de semblances entre documents amb el mètode de TF-IDF:

Per calcular les semblances entre dos documents amb el mètode TF-IDF, es representen com a vectors el conjunt de paraules no funcionals amb la mateixa estructura que al càlcul de similituds pel mètode del cosinus però amb una diferent interpretació dels pesos dels termes que es calculen amb la següent fórmula:

$$\log\left(\frac{d}{dfj}\right)$$

On d és el nombre total de documents i dfj és el nombre de documents que contenen aquest terme del qual es calcula el seu pes.

Per calcular la similitud entre dos documents és fa el producte escalar ($\vec{v} \times \vec{w}$) i el resultat dóna un nombre entre 0 i 1 que és el grau de similitud que tenen el dos documents.

4. Llista de Classes Implementades

| Classe | Autor |
|-------------------|------------------------|
| Archivo | John Bryan Game Glices |
| Binario | John Bryan Game Glices |
| Cjt_Autors | Albert Ruiz Lombarte |
| Contingut | Aleix Roselló Rosiñol |
| CtrlDomini | Albert Ruiz Lombarte |
| CtrlPresPrincipal | Albert Ruiz Lombarte |
| Document | Aleix Roselló Rosiñol |
| ExpresionTree | Aleix Roselló Rosiñol |
| Frase | Aleix Roselló Rosiñol |
| GestorArchivo | John Bryan Game Glices |
| GestorCjt_Autors | Albert Ruiz Lombarte |
| GestorDocuments | Aleix Roselló Rosiñol |
| GestorMatriu | Albert Ruiz Lombarte |
| Matriu | Albert Ruiz Lombarte |
| Node | Aleix Roselló Rosiñol |
| Pair | Albert Ruiz Lombarte |
| PairComparator | Albert Ruiz Lombarte |
| Pres_AddDoc | Albert Ruiz Lombarte |
| Pres_AP | Albert Ruiz Lombarte |
| Pres_ConDoc | Albert Ruiz Lombarte |
| Pres_DelDoc | Albert Ruiz Lombarte |
| Pres_DS | Albert Ruiz Lombarte |
| Pres_EB | Albert Ruiz Lombarte |
| Pres_ExpDoc | Albert Ruiz Lombarte |
| Pres_ImpDoc | Albert Ruiz Lombarte |
| Pres_ImpPF | Albert Ruiz Lombarte |
| Pres_Llista | Albert Ruiz Lombarte |
| Pres_ModDoc | Albert Ruiz Lombarte |
| Pres_TA | Albert Ruiz Lombarte |

5. Relació de Llibreries externes utilitzades

Les úniques llibreries externes que hem utilitzat són les següents:

1. Java.util.*

Concremanent: Java.util.ArrayList, Java.util.Collections, Java.util.Comparators i Java.util.PriorityQueue

Utilitzada per gestionar llistes anomenades ArrayList, per a l'us d'algorismes de cerca dicotòmica i ordenació a més de comparadors per organitzades en cues de prioritat.

2. Java.io.*

Concretament: Java.io.BufferedReader, Java.io.BufferedWriter, Java.io.IOException, Java.io.InputStream

Utilitzada per gestionar les dades d'entrada i de sortida del programa ja sigui via fitxers de text o fitxers binaris, així com les excepcions que poden generar.

3. Javax.swing.*

Concretament: javax.swing.AbstractAction, javax.swing.Action, javax.swing.ImageIcon, javax.swing.JdesktopPane, javax.swing.JFileChooser, javax.swing.JFrame, javax.swing.JinternalFrame, javax.swing.Jmenu, javax.swing.JmenuBar, javax.swing.JoptionPane.

Utilitzada per generar i gestionar els components de la presentació ja siguin panells, botons, quadres de text, llistats, navegadors de fitxers, etc...

6. Llistat de funcionalitats implementades

1. Gestió de Sessió:

Descripció: Permet a l'usuari navegar les funcionalitats relacionades amb la gestió de la seva sessió.

Comportament: L'usuari pot seleccionar d'entre les opcions la funcionalitat de gestió de la sessió que vol utilitzar.

2. Obrir Sessió:

Descripció: Permet a l'usuari obrir la informació d'una sessió que ja havia desat anteriorment.

Comportament: L'usuari selecciona el fitxer binari amb el navegador de fitxers en el directori on l'havia desat.

3. Guardar Sessió:

Descripció: Permet a l'usuari desar la informació de la sessió actual.

Comportament: L'usuari indica el nom i el directori amb el navegador de fitxers de la sessió que vol desar.

4. Sortir:

Descripció: Permet a l'usuari finalitzar el programa.

Comportament: El sistema acaba amb el procés que estava executant.

5. Gestió de dades:

Descripció: Permet a l'usuari navegar per les funcionalitats relacionades amb la gestió de les dades de les que disposa el sistema.

Comportament: L'usuari pot seleccionar d'entre les opcions la funcionalitat de gestió que vol utilitzar.

6. Afegir Document:

Descripció: Permet a l'usuari afegir un nou document al sistema.

Comportament: L'usuari indica el títol, l'autor i el contingut del nou document que vol afegir, si l'autor no existeix s'afegirà amb la seva nova obra, si existeix s'afegirà a l'autor l'obra que ha escrit, i si l'autor ja ha escrit una obra amb aquest títol s'indicarà a l'usuari un error.

7. Modificar Document:

Descripció: Permet a l'usuari modificar qualsevol document del sistema.

Comportament: L'usuari indica el títol i l'autor del document al que vol modificar el seu contingut, si l'autor no existeix o no ha escrit una obra amb aquest títol s'indicarà a l'usuari un error, sinó, podrà modificar els seu contingut satisfactòriament.

8. Esborrar Document:

Descripció: Permet a l'usuari esborrar permanentment qualsevol document del sistema.

Comportament: L'usuari indica el títol i l'autor del document que vol esborrar del sistema, si l'autor no existeix o no ha escrit una obra amb aquest títol s'indicarà a l'usuari un error, sinó, el document s'esborrarà del sistema satisfactòriament.

9. Importar Paraules Funcionals:

Descripció: Permet a l'usuari importar un fitxer que contingui paraules funcionals per a que no es tinguin en compte a l'hora de calcular les similituds entre documents.

Comportament: L'usuari indica la ruta absoluta i el nom del fitxer que vol importar, si la ruta o nom no són vàlids s'indicarà una excepció. Si és correcta, s'afegiran les paraules funcionals a la llista de paraules a ignorar en calcular similituds entre documents.

10. Importar Document:

Descripció: Permet a l'usuari importar un fitxer que contingui la informació d'un document.

Comportament: L'usuari indica el títol, l'autor i amb el navegador de fitxers ruta absoluta i el nom del fitxer amb el contingut que vol importar, si la ruta o nom no són vàlids s'indicarà un error . Si és correcta. Si l'autor no existeix s'afegirà amb la seva nova obra, si existeix s'afegirà a l'autor l'obra que ha escrit, i si l'autor ja ha escrit una obra amb aquest títol s'indicarà a l'usuari un error.

11. Exportar Document:

Descripció: Permet a l'usuari exportar un fitxer que contingui la informació d'un document.

Comportament: L'usuari indica la ruta absoluta i el nom del fitxer que vol exportar amb el navegador de fitxers, si la ruta o nom no són vàlids s'indicarà una excepció. Si és correcta el contingut d'aquell document es desarà en el fitxer de text.

12. Consultar Dades:

Descripció: Permet a l'usuari navegar per les funcionalitats relacionades amb les consultes de les dades del sistema.

Comportament: L'usuari pot seleccionar d'entre les opcions la funcionalitat de consulta que vol utilitzar.

13. Consultar Títols d'un Autor:

Descripció: Permet a l'usuari consultar els títols de les obres que ha escrit un determinat autor.

Comportament: L'usuari indica el nom de l'autor del qual vol consultar els títols de les seves obres, si l'autor no existeix se l'indicarà a l'usuari amb un error.

14. Consultar Autors per prefix:

Descripció: Permet a l'usuari consultar els autors del sistema mitjançant un prefix.

Comportament: L'usuari indica el prefix, que pot ser buit, i se li mostraran els noms dels autors que comencin per aquest prefix.

15. Consultar Llistat Documents:

Descripció: Permet a l'usuari consultar la llista de documents que hi ha al sistema.

Comportament: A l'usuari se li mostra una llista de tots els documents identificats per autor i títol que hi ha al sistema.

16. Consultar Contingut:

Descripció: Permet a l'usuari consultar el contingut d'un determinat document.

Comportament: L'usuari indica el títol i l'autor del document del que vol consultar el seu contingut, si l'autor no existeix o bé no ha escrit una obra amb aquest títol se l'indicarà a l'usuari amb un error.

17. Consultar Documents similars:

Descripció: Permet a l'usuari consultar el subconjunt de documents més similars a un donat.

Comportament: L'usuari indica el títol i l'autor del document del que vol consultar el subconjunt de documents més semblants, si l'autor no existeix o bé no ha escrit una obra amb aquest títol se l'indicarà a l'usuari amb un error. Després se li demanarà a l'usuari el mètode de càlcul de similituds que vol usar i un nombre natural k , i se li mostraran els k documents, títol, autor i grau de semblança, ordenats de més similar a menys similar al que ha indicat.

18. Consultar Documents expressió booleana:

Descripció: Permet a l'usuari consultar el subconjunt de documents els contingut dels quals compleixi una determinada expressió booleana.

Comportament: L'usuari indica una expressió booleana i se li mostren els títols i autors de les obres el contingut de les quals compleixen l'expressió. Però si no està en el format correcte se l'indicarà a l'usuari amb un error.

19. Ajuda:

Descripció: Permet a l'usuari navegar per les funcionalitats d'ajuda.

Comportament: L'usuari pot seleccionar d'entre les opcions la funcionalitat d'ajuda que vol utilitzar.

20. Manual d'usuari:

Descripció: Permet a l'usuari consultar el manual d'usuari del programa

Comportament: S'obrirà un visualitzador de fitxers pdf per a que l'usuari pugui veure el manual.

21. About:

Descripció: Permet a l'usuari consultar l'autoria del software.

Comportament: S'obrirà un missatge d'avís amb l'informació de l'autoria del programa.

7. Manual d'usuari

Manual d'usuari

Projecte de Programació Subgrup 22 clúster 6.2

Gestió de la Sessió:

Permet navegar per les funcionalitats relacionades amb la gestió de la sessió que conté tots els canvis en les dades desades fins al moment. Existeixen les següents funcionalitats que es presenten a continuació:



Obrir Sessió:

Permet obrir el fitxer binari amb les dades d'una sessió desada anteriormemt, per seleccionar-lo cerca el fitxer binari en el navegador de fitxers un cop trobat, clica sobre el fitxer i despres a Open.



Guardar Sessió:

Permet desar en un fitxer binari les dades dels canvis que s'han produït en la sessió actual. Per crear el fitxer cal seleccionar el directori amb el navegador de fitxers dins del qual es crearà el binari amb les dades desades. Un cop seleccionat el directori, s'anomena el fitxer i es prem al botó de Save.



Sortir:

Permet tancar el programa que s'estava executant, si no s'han desat els canvis en les dades amb la funcionalitat de Guardar Sessió es perdran irreversiblement.



Gestió de dades:

Permet navegar per les funcionalitats relacionades amb la gestió de les dades de les que disposa el sistema. Existeixen les següents funcionalitats per gestionar les dades:



Afegir Document:

Permet afegir un document, introduint el títol, l'autor i el contingut en els diferents camps en blanc, no es podrà afegir cap document que contingui qualsevol dels camps buit.



Esborrar Document:

Permet esborrar un document de les dades de les que disposa el sistema indicant els seu títol i autor.

**Modificar Document:**

Permet modificar el contingut d'un document. Per fer-ho cal indicar el títol i l'autor del document que es vol modificar, premer el botó de mostrar el contingut que indicarà el contingut actual del document per a ser modificat, un cop modificat cal premer el botó de modificar document per desar els canvis realitzats.

**Importar Document:**

Permet importar el contingut d'un document que es troba en un fitxer de text, per fer-ho cal indicar el títol i el nom del seu autor, després cal seleccionar el fitxer de text amb el navegador de fitxers i seleccionar Open. Un cop fet això el sistema disposarà d'un nou document amb el títol i autor indicats i amb contingut igual al text que contenia el fitxer seleccionat.

**Importar Paraules Funcionals:**

Permet importar un fitxer amb noves paraules funcionals que no es tindran en compte en el càlcul de similituds entre documents. Per fer-ho cal seleccionar el fitxer de text amb el navegador de fitxers i premer el botó de Open. Un cop fet això el sistema interpretarà cada línia del fitxer de text com una paraula funcional que no es tindrà en compte en el càlcul de similituds entre documents.

**Exportar Document:**

Permet exportar el contingut d'un document, per fer-ho cal indicar el títol i l'autor del document i indicar el nom del fitxer en el que s'exportarà el contingut del document i premer el botó de Save.

**Consultes de les dades del sistema:**

Permet navegar per les funcionalitats relacionades amb les consultes de les dades del sistema. Existeixen les següents funcionalitats per cerca dades en el sistema:

**Consultar Contingut:**

Permet consultar el contingut d'un document tot indicant el seu títol i el seu autor. Per fer-ho cal indicar el títol i l'autor i premer el botó de mostrar contingut.

**Llistar títols d'un autor:**

Permet consultar els títols dels documents que ha escrit un determinat autor tot indicant al sistema el nom d'aquest. Per fer-ho cal indicar el nom de l'autor i premer el botó de llistar títols.

**Llistar autors per prefix:**

Permet consultar els noms dels autors que comencen per un determinat prefix indicat per l'usuari. Per fer-ho cal indicar el prefix que pot ser buit i premer el botó de llistar autors.

**Llistar documents:**

Permet consultar els títols i autors de tots els documents que hi han al sistema.

**Llistar documets similars:**

Permet consultar els N documents més similars a un donat amb dos mètodes de càlcul de similituds de continguts (Cosinus i TF-IDFJ). Per fer-ho l'usuari ha d'indicar el títol i l'autor del document amb el que vol cercar els seus similars , el nombre de documents que vol veure, seleccionar el mètode de càlcul de similituds que vol usar i premer el botó de llistar documents similars. Aleshores se li mostraran els N documents més similars al donat per ordre de semblança i amb el grau de similitud indicat amb un nombre entre 0 i 1.

**Llistar documents per expressió booleana:**

Permet consultar els documents que compleixen l'expressió booleana indicada. Per fer-ho cal indicar la expressió booleana ben parentitzada i premer el botó de cercar documents.

**Ajuda:**

Permet navegar per les funcionalitats d'ajuda a l'usuari i informació del programa, existeixen les següents funcionalitats relacionades amb l'ajuda a l'usuari i informació:

**Manual d'usuari:**

Permet consultar aquest document en format pdf amb les funcionalitats del programa.

**About:**

Permet consultar l'autoria del software.

8. Jocs de prova complets

Prova 1

Descripció: Joc de prova que testeja les següents funcionalitats en el següent ordre:

1. Importar Document, 2. Consultar Documents, 3. Consultar Similitud entre Documents amb el mètode de càlcul de TF-IDF.

La prova consistirà en el següent:

1.- Importar els documents que tenen el següent contingut:

D1: el río Danubio pasa por Viena, su color es azul

D2: el caudal de un río asciende en Invierno

D3: el río Rhin y el río Danubio tienen mucho caudal

D4: si un río es navegable, es porque tiene mucho caudal

Q: cuál es el caudal del río Danubio

2.- Consultar que els documents realment formen part del sistema

3.- Consultar la similitud dels documents Q amb tots els demés per veure si realment el càlcul que és mostra a continuació és correcte:

Tenim la següent matriu on els pesos són el nombre d'aparicions de cada paraula:

| | Río | Danubio | Viena | color | azul | caudal | invierno | Rhin | navegable |
|----|-----|---------|-------|-------|------|--------|----------|------|-----------|
| D1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 |
| D2 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 |
| D3 | 2 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 0 |
| D4 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 |

Càlcul de freqüències inverses:

$\text{Idf}(\text{río}) = \text{Log}(4/4) = \log(1) = 0$

$\text{Idf}(\text{Danubio}) = \text{Log}(4/2) = \log 2 = 0.301$ $\text{Idf}(\text{Viena}) = \text{Log}(4/1) = \log 4 = 0.602$ $\text{Idf}(\text{color}) = \text{Log}(4/1) = \log 4 = 0.602$

$\text{Idf}(\text{azul}) = \text{Log}(4/1) = \log 4 = 0.602$

$\text{Idf}(\text{caudal}) = \text{Log}(4/3) = \log 1.33 = 0.124$ $\text{Idf}(\text{invierno}) = \text{Log}(4/1) = \log 4 = 0.602$ $\text{Idf}(\text{Rhin}) = \text{Log}(4/1) = \log 4 = 0.602$

$\text{Idf}(\text{navegable}) = \text{Log}(4/1) = \log 4 = 0.602$

Queda la següent matriu TF-IDF:

| | Río | Danubio | Viena | color | azul | Caudal | invierno | Rhin | navegable |
|----|-----|---------|-------|-------|-------|--------|----------|-------|-----------|
| D1 | 0 | 0.301 | 0.602 | 0.602 | 0.602 | 0 | 0 | 0 | 0 |
| D2 | 0 | 0 | 0 | 0 | 0 | 0.124 | 0.602 | 0 | 0 |
| D3 | 0 | 0.301 | 0 | 0 | 0 | 0.124 | 0 | 0.602 | 0 |
| D4 | 0 | 0 | 0 | 0 | 0 | 0.124 | 0 | 0 | 0.602 |
| Q | 0 | 0.301 | 0 | 0 | 0 | 0.124 | 0 | 0 | 0 |

Càlcul de similituds (producte escalar dels vectors):

$\text{Sim}(D1, Q) = 0*0 + 0.301 * 0.301 + 0.602*0 + 0.602*0 + 0.602*0 + 0*0.124 + 0*0 + 0*0 + 0*0 = 0.090601...$

$\text{Sim}(D2, Q) = 0*0 + 0*0.301 + 0*0 + 0*0 + 0*0 + 0.124*0.124 + 0.602*0 + 0*0 + 0*0 = 0.0153376$

$\text{Sim}(D3, Q) = 0*0 + 0.301 * 0.301 + 0*0 + 0*0 + 0*0 + 0.124*0.124 + 0*0 + 0.602*0 + 0*0 = 0.105977$

$\text{Sim}(D4, Q) = 0*0 + 0*0.301 + 0*0 + 0*0 + 0*0 + 0.124*0.124 + 0*0 + 0*0 + 0.602*0 = 0.015337$

Aleshores tenim com a resultat de consultar la similitud de Q amb els altres documents del sistema la següent llista ordenada de més similar a menys:

D3 (0.105977)

D1 (0.090601)

D2 (0.015337)

D4 (0.015337)

Objectius: L'objectiu principal és comprobar el bon funcionament del mètode de càlcul de TF-IDF a part de l'ús de les funcionalitats necessàries per ús d'aquesta funcionalitat principal que es vol testejar.

Entrada:

Gestió de dades > Importar Document

Títol: D1

Autor: nom_autor

Navegador de fitxers: EXE/data/Prova1/D1.txt

Open

Títol: D2

Autor: nom_autor

Navegador de fitxers: EXE/data/Prova1/D2.txt

Open

Títol: D3

Autor: nom_autor

Navegador de fitxers: EXE/data/Prova1/D3.txt

Open

Títol: D4

Autor: nom_autor

Navegador de fitxers: EXE/data/Prova1/D4.txt

Open

Títol: Q

Autor: nom_autor

Navegador de fitxers: EXE/data/Prova1/Q.txt

Open

Tanca

Cerca > Llistar Documents

Sortida:

D1 – nom_autor
D2 – nom_autor
D3 – nom_autor
D4 – nom_autor
Q – nom_autor

Entrada:

Tanca
Cerca > Llistar documents similars
Títol: Q
Autor: nom_autor
Quin mètode de càlcul vols usar? Càlcul amb el mètode TF-IDFJ
Quin nombre de documents vols veure? 4

Sortida:

1. D3 – nom_autor : 0.105977
2. D1 – nom_autor : 0.090601
3. D2 – nom_autor : 0.015337
4. D4 – nom_autor : 0.015337

Prova 2:

Descripció: Joc de prova complet que prova les següents funcionalitats en el següent ordre:

1. Importar Document, 2. Llistar Documents, 3. Consultar Autors per prefix, 4. Guardar Sessió 5. Sortir 6. Obrir Sessió 7. Llistar Documents, 8. Consultar Contingut.

Entrada:

Gestió de dades > Importar Document
Títol: L'art de (no) callar
Autor: Esther Vera
Navegador de fitxers: EXE/data/Prova2/article1.txt
Títol: Realitats i percepcions
Autor: Carme Colomina
Navegador de fitxers: EXE/data/Prova2/article2.txt
Títol: A vegades ens en sortim, Sebastià
Autor: Carles Capdevila
Navegador de fitxers: EXE/data/Prova2/article3.txt
Títol: La fi del món
Autor: Fernando Trias de Bes
Navegador de fitxers: EXE/data/Prova2/article4.txt
Títol: Estar segur de saber
Autor: Joaquim Coello
Navegador de fitxers: EXE/data/Prova2/article5.txt
Cerca > Llistar Documents

Sortida:

L'art de (no) callar - Esther Vera
Realitats i percepcions - Carme Colomina
A vegades ens en sortim, Sebastià - Carles Capdevila
La fi del món - Fernando Trias de Bes
Estar segur de saber - Joaquim Coello

Entrada:

Cerca > Llistar Autors per prefix
Prefix: Car
Llistar autors per prefix

Sortida:

Carles Capdevila
Carme Colomina

Entrada:

Sessió > Guardar Sessió
Navegador de fitxres: EXE/data/Prova2
File Name: prova2
Sessió > Sortir
Obrir el programa de nou
Sessió > Obrir Sessió
Navegador de fitxres: EXE/data/Prova2/prova2
Cerca > Llistar documents

Sortida:

L'art de (no) callar - Esther Vera
Realitats i percepcions - Carme Colomina
A vegades ens en sortim, Sebastià - Carles Capdevila
La fi del món - Fernando Trias de Bes
Estar segur de saber - Joaquim Coello

Entrada:

Cerca > Consultar Contingut
Títol: Realitats i percepcions
Autor: Carme Colomina
Cercar Contingut

Sortida:

(Contingut del document)

Prova 3

Descripció: Joc de prova que serveix per provar la funcionalitat de consulta per expressió booleana.

La prova consisteix del següent:

- 1- Importar els diferents articles que es troben a la carpeta EXE/data/Prova2/
- 2- Realitzem la següent consulta booleana:
((silenci & cultura) | ("principis de desembre" & !silenci))
- 3- Comprovar que els documents compleixen la condició a partir del contingut

Importem amb:

Gestió de dades > Importar document
Introduim les dades
Títol: t0
Autor: a0
així succesivament...

A continuació fem la consulta amb:

Cerca > Llistar documents per expressió booleana

Introduim la següent expressió:
((cultura) | ("principis de desembre" & !silenci))

Veiem que la sortida és:

t0 - a0

t1 - a1

Que efectivament són els documents que compleixen l'expressió booleana.